Prioritizing Amyloid Imaging Biomarkers in Alzheimer's Disease via Learning to Rank

Bo Peng[1], Zhiyun Ren[1], Xiaohui Yao[2], Kefei Liu[2], Andrew J. Saykin[3], Li Shen[2], Xia Ning[1], and *for the ADNI*[⋆]

[1] The Ohio State University, Columbus OH 43210, USA
peng.707@buckeyemail.osu.edu
ren.685@osu.edu
ning.104@osu.edu
[2] University of Pennsylvania, Philadelphia PA 19104, USA
xiaohui.yao@pennmedicine.upenn.edu
kefei.liu@pennmedicine.upenn.edu
li.shen@pennmedicine.upenn.edu
[3] Indiana University School of Medicine, Indianapolis IN 46202, USA
asaykin@iupui.edu

Abstract. We propose an innovative machine learning paradigm enabling precision medicine for AD biomarker discovery. The paradigm tailors the imaging biomarker discovery process to individual characteristics of a given patient. We implement this paradigm using a newly developed learning-to-rank method `PLTR`. The `PLTR` model seamlessly integrates two objectives for joint optimization: pushing up relevant biomarkers and ranking among relevant biomarkers. The empirical study of `PLTR` conducted on the ADNI data yields promising results to identify and prioritize individual-specific amyloid imaging biomarkers based on the individual's structural MRI data. The resulting top ranked imaging biomarker has the potential to aid personalized diagnosis and disease subtyping.

Keywords: Amyloid PET · Structural MRI · Imaging biomarker prioritization · Learning to rank · Alzheimer's disease.

1 Introduction

Alzheimer's disease (AD) is a national priority, with 5.5 million Americans affected at an annual cost of $259 billion in 2017 and no available cure [1]. Brain characteristics related to AD progression may be captured by multimodal magnetic resonance imaging (MRI) and positron emission tomography (PET) scans.

⋆ Data used in preparation of this article were obtained from the Alzheimer¡s Disease Neuroimaging Initiative (ADNI) database (ad-ni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Thus, there is a large body of neuroimaging studies in AD, aiming to develop image-based predictive machine learning models for early detection of AD as well as identification of relevant imaging biomarkers (e.g., [8]). These models are typically designed to accomplish learning tasks such as regression, classification and/or survival analysis. As a result, the identified imaging biomarkers are at the population level and not specific to an individual subject.

In this work, we propose a novel learning paradigm that embraces the concept of precision medicine and tailors the imaging biomarker discovery process to the individual characteristics of a given patient. Specifically, we perform an innovative application of a newly developed learning-to-rank method, denoted as `PLTR` [5], to the structural MRI and amyloid PET data of the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort [9]. Using structural MRI as the individual characteristics, our goal is to not only identify individual-specific amyloid imaging biomarkers but also prioritize them according to AD-specific abnormality. Compared with traditional biomarker studies at the population level, the uniqueness of our study is twofold: (1) the identified biomarkers are tailored to each individual patient; and (2) the identified biomarkers are prioritized based on the individual's characteristics, which has the potential to enable personalized diagnosis and disease subtyping.

2 Materials and Data Processing

To demonstrate the effectiveness of the learning-to-rank method for personalized prioritization of the amyloid imaging biomarkers, we applied it to the ADNI cohort [9]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI, a prodromal stage of AD) and early AD. For up-to-date information, see www.adni-info.org.

Data used in the preparation of this article were obtained from the 2017 ADNI TADPOLE grand challenge (tadpole.grand-challenge.org/), and was downloaded from the ADNI website (adni.loni.usc.edu). The TADPOLE data used in this study consists of structural MRI and AV45-PET (amyloid) imaging data as well as diagnostic information. Both MRI and amyloid imaging data have been pre-processed with standard ADNI pipelines as described previously in [7].

In this study, we included all the regional MRI measures with field name containing "UCSFFSX" in the TADPOLE D1 and D2 data sets. Specifically, these are FreeSurfer regional volume and cortical thickness measures processed by the ADNI UCSF team. We also included all the regional amyloid measures with field name containing "UCBERKELEYAV45" in the TADPOLE D1 and D2 data sets. These are cortical and subcortical amyloid deposition measures processed by the ADNI UC Berkeley team.

Originally, there are totally 12,741 participant visit records with 103 amyloid features, 125 FreeSurfer features and diagnostic information corresponding to

each visit. To convert this longitudinal data set into a cross-sectional one as well as handle the missing data issue, we use the following procedure to generate a clean set of cross-sectional data: (1) remove visit records that have more than 50 percent of null values across 103 amyloid features, with 10,623 records removed; (2) extract the earliest AV45-PET visit for each participant, with 1,091 records kept; (3) remove visit records that have more than 50 percent of null values across 125 FreeSurfer features, with 58 records removed; (4) remove features that have more than 50 percent of null values across records, with 16 FreeSurfer features removed; (5) remove 3 participants with no diagnostic information. Finally, 1,030 participants with 103 amyloid and 109 FreeSurfer measures are studied, including 351 health control (HC), 501 MCI and 178 AD participants. We treat both MCI and AD subjects as patients, and so have a total of 679 cases and 351 controls.

3 Methods

We use the joint push and learning-to-rank method as developed in He *et al.* [5], denoted as `PLTR`, for personalized patient feature prioritization. Our goal is to prioritize amyloid features for each patient that are most relevant to his/her disease diagnosis using patients' existing information. The underlying hypothesis is that patients with similar FreeSurfer feature profiles would have similar ranking structures among their amyloid features. In the context of AD feature prioritization, `PLTR` learns and uses latent vectors of patients and amyloid features to score each amyloid feature for each patient, and ranks the features based on their scores; patients with similar FreeSurfer feature profiles will have similar latent vectors. During the learning process, `PLTR` explicitly pushes the most relevant amyloid features on top of the less relevant ones for each patient, and therefrom optimizes the latent patient and amyloid feature vectors so they will reproduce the pushed ranking structures.

3.1 Overview of `PLTR`

In `PLTR`, the ranking of features in terms of their relatedness to MCI/AD in a patient is determined by their latent scores on the patient. For a feature f_i and a patient $\mathcal{P}_p$, f_i's latent score on $\mathcal{P}_p$, denoted as $s_p(f_i)$, is calculated as the dot product of f_i's latent vector $\mathbf{v}_i \in \mathbb{R}^{l\times 1}$ and $\mathcal{P}_p$'s latent vector $\mathbf{u}_p \in \mathbb{R}^{l\times 1}$, where l is the latent dimension, as follows,

$$s_p(f_i) = \mathbf{u}_p^\mathsf{T}\mathbf{v}_i, \tag{1}$$

where the latent vectors $\mathbf{u}_p$ and $\mathbf{v}_i$ will be learned. All the features are then sorted based on their scores on $\mathcal{P}_p$, with the most relevant features having the highest scores and ranked higher than irrelevant features.

Overall, `PLTR` seeks the patient latent vectors and feature latent vectors that will be used in feature scoring function s (Equation (1)) such that for each patient, the relevant features will be ranked on top and in right orders using the

latent vectors. In PLTR, such latent vectors are learned by solving the following optimization problem:

$$\min_{U,V} \mathcal{L}_s = (1-\alpha)P_s^{\uparrow} + \alpha O_s^{+} + \frac{\beta}{2}R_{uv} + \frac{\gamma}{2}R_{\text{csim}}, \tag{2}$$

where $\mathcal{L}_s$ is the overall loss function; $P_s^{\uparrow}$ measures the number of irrelevant features that are ranked on top of relevant features; O_s^{+} measures the ranking among relevant features. R_{uv} is a regularizer on U and V to prevent overfitting, defined as

$$R_{uv} = \frac{1}{m}\|U\|_F^2 + \frac{1}{n}\|V\|_F^2, \tag{3}$$

where m and n are the number of patients and the number of features, respectively; $\|X\|_F$ is the Frobenius norm of matrix X. R_{csim} is a regularizer on patients to constrain patient latent vectors, defined as

$$R_{\text{csim}} = \frac{1}{m^2}\sum_{p=1}^{m}\sum_{q=1}^{m} w_{pq}\|\mathbf{u}_p - \mathbf{u}_q\|_2^2, \tag{4}$$

where w_{pq} is the similarity between $\mathcal{P}_p$ and $\mathcal{P}_q$ that is calculated using FreeSurfer features of the patients. Details of these terms can be found in He *et al.* [5].

3.2 Patient Similarities from FreeSurfer Features

We consider 109 FreeSurfer features and represent each patient as a FreeSurfer feature vector, denoted as $\mathbf{r}_p = [r_{p1}, r_{p2}, \cdots, r_{pn_r}]$, where r_{pi} $(i = 1, \cdots, n_r)$ is a FreeSurfer feature for patient p. Thus, for all the patients, we construct a FreeSurfer feature matrix $R_{\text{AD}} = [\mathbf{r}_1^{+}; \mathbf{r}_2^{+}; \cdots; \mathbf{r}_{m^+}^{+}] \in \mathbb{R}^{m^+ \times n_r}$ and for all the health control subjects (HCs), a FreeSurfer feature matrix $R_{\text{HC}} = [\mathbf{r}_1^{-}; \mathbf{r}_2^{-}; \cdots; \mathbf{r}_{m^-}^{-}] \in \mathbb{R}^{m^- \times n_r}$, where m^+ and m^- are the numbers of AD/MCI patients and HCs, respectively, and n_r is the number of FreeSurfer features. We scale R_{AD} values into the unit interval by dividing each column of R_{AD} (i.e., each FreeSurfer feature) using its maximum value. The normalized R_{AD} matrix is denoted as $\bar{R}_{\text{AD}}$, and the similarities between patients are calculated over $\bar{R}_{\text{AD}}$ using the radial basis function (RBF) kernel:

$$w_{pq} = \exp\left(-\frac{\|\bar{R}_{\text{AD}}(p,:) - \bar{R}_{\text{AD}}(q,:)\|^2}{2\sigma^2}\right), \tag{5}$$

where w_{pq} is the patient similarity used in Equation (4). This patient similarity measurement is denoted as `simU`.

3.3 Patient Amyloid Features in Ground Truth

Similarly, each patient is also represented by an amyloid feature vector, denoted as $\mathbf{c}_p = [c_{p1}, c_{p2}, \cdots, c_{pn_c}]$, where c_{pi} $(i = 1, \cdots, n_c)$ is an amyloid feature for patient p. Thus, we construct an amyloid feature matrix $C_{\text{AD}} = [\mathbf{c}_1^{+}; \mathbf{c}_2^{+}; \cdots, \mathbf{c}_{m^+}^{+}]$

for AD/MCI patients, and an amyloid feature matrix $C_{\mathrm{HC}} = [\mathbf{c}_1^-; \mathbf{c}_2^-; \cdots, \mathbf{c}_{m^-}^-]$ for HC subjects. We normalize C_{AD} by dividing each column of C_{AD} (i.e., each amyloid feature) by the mean value of the corresponding column in C_{HC}. Thus, the normalization results in C_{AD} measure the extent to which an amyloid feature in patients deviates from that in HCs. The normalized matrix, denoted as $\bar{C}_{\mathrm{AD}}$, is used as the ground truth of amyloid feature ranking. That is, the optimization problem (2) tries to learn the latent vectors that reconstruct the ordering structures in $\bar{C}_{\mathrm{AD}}$, and prioritize amyloid features that are most relevant to patients. The reason why we use FreeSurfer features to quantitatively measure patients and prioritize amyloid features correspondingly is that MRI imaging is non-invasive and relatively low-cost as compared to PET imaging.

4 Experiments

4.1 Experimental Protocol

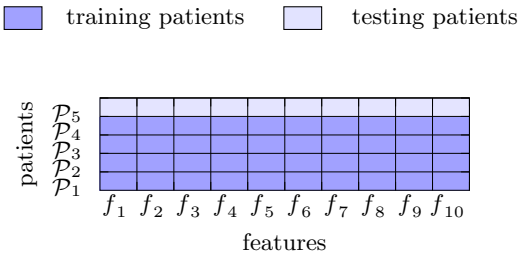

Fig. 1: Data split

We split patients into training set and testing set, such that a certain patient and all his/her features will be either in the training set or in the testing set. We train the `PLTR` model using training patients and test its performance on the testing patients. This corresponds to the use scenario in which we want to identify the most potentially useful AD biomarkers for new patients, based on the existing information of the patients, when such biomarkers have not been tested on the new patients. Figure 1 demonstrates the data split process.

We define average hit at k, denoted as AH@k, to evaluate the ranking performance. AH@k is defined as follows:

$$\mathrm{AH@}k(\tau, \tilde{\tau}) = \sum_{i=1}^{k} \mathbb{I}(\tilde{\tau}_i \in \tau), \tag{6}$$

where τ is the ground-truth ranking list, $\tilde{\tau}$ is the predicted ranking list, and $\tilde{\tau}_i$ is the i-th ranked item in $\tilde{\tau}$. That is, AH@k calculates the number of items among top k in the predicted lists that are also in the ground truth (i.e., hits). Higher AH@k values indicate better prioritization performance.

We define a second evaluation metric weighted average hit at k, denoted as WAH@k as follows:

$$\text{WAH@}k(\tau, \tilde{\tau}) = \sum_{j=1}^{k} \sum_{i=1}^{j} \mathbb{I}(\tilde{\tau}_i \in \tau)/j, \tag{7}$$

that is, WAH@k is a weighted version of AH@k that calculates the average hit up to top k. Higher WAH@k indicate more hits and those hits are ranked on top in the ranking list. By default, the ground-truth τ has k items (i.e., the top-k items among all the sorted items) in Equation (6) and Equation (7).

4.2 Baseline Methods

We compare `PLTR` with another two methods: the Bayesian Multi-Task Multi-Kernel Learning (`BMTMKL`) method [2] and the Kernelized Rank Learning (`KRL`) method [4]. `BMTMKL` uses kernels over FreeSurfer features to predict amyloid feature values. `KRL` uses kernel regression with a ranking loss to predict amyloid feature values. These two methods represent two strong baseline methods for the biomarker feature prioritization problem. We use the patient similarity matrix 5 as the kernels for `BMTMKL` and `KRL`. We conducted parameter grid search to identify the best parameters for each model, and present the best performance of the models.

5 Experimental Results

5.1 Overall Performance

We first hold out 35 and 163 patients as testing patients, respectively. These testing patients are determined such that they have more than 10 similar patients in the training set, and the corresponding patient similarities are higher than 0.75 and 0.65, respectively. Patient latent vectors and feature latent vectors are learned on the training patients. The feature scores for the testing patients are calculated as the weighted sum of the predicted feature scores from their top-10 most similar training patients, where the weights are the corresponding patient similarities. The patient similarities are calculated using `simU` (Equation (5), $\sigma = 1$). The patient amyloid features are normalized as described in Section 3.3. Please note that we only use patients (i.e., MCI and AD subjects) for model training and testing, and only use controls (i.e., HC subjects) to set the standard for patient data normalization, as feature prioritization for healthy controls has limited clinical interests.

Table 1 presents the best performance of `PLTR` in terms of AH@5 for each latent dimension. When 35 patients are hold out for testing, the best AH@5 is 1.886 when latent dimension $d = 20$, and the corresponding WAH@5 is 1.632. This performance is significantly better than those of the baseline methods. Note that we use predicted feature scores to prioritize features for the testing patients.

Table 1: Overall Performance of PLTR (simU, $\sigma = 1$)

n	method	Parameters					AH@5	WAH@5	AH@10	WAH@10
		α	β	γ	d	λ				
35	PLTR	0.3	0.5	1.0	10	-	1.857	1.545	**3.371**	2.249
		0.3	0.5	1.0	20	-	**1.886**	**1.632**	3.286	1.987
		0.3	0.5	1.0	50	-	1.857	1.560	3.314	2.007
	BMTMKL	-	-	-	-	-	0.971	0.916	2.171	**2.573**
	KRL	-	-	-	-	3.0	0.429	0.426	1.086	1.245
163	PLTR	0.5	1.0	1.0	10	-	1.343	0.930	3.080	**2.497**
		0.5	1.0	1.0	20	-	**1.429**	**1.067**	3.074	2.402
		0.5	1.0	1.0	50	-	**1.429**	1.012	**3.110**	2.437
	BMTMKL	-	-	-	-	-	0.282	0.288	0.957	0.929
	KRL	-	-	-	-	0.1	0.356	0.389	1.025	1.054

The column "n" corresponds to the number of hold-out testing patients. Best performance under each evaluation metric is in **bold**.

Table 1 also shows that PLTR significantly outperforms the baseline methods in terms of AH@10. PLTR is slightly worse than BMTMKL on WAH@10 (2.249 vs 2.573). This indicates that among top 10 drugs in the ranking list, PLTR is able to rank more relevant features on top than BMTMKL, although the positions of those hits are not as high as BMTMKL. When 163 patients are hold out for testing, the best performance of PLTR (i.e., AH@5 1.429 when $d = 20$) is still better than those of the baseline methods. This indicates that PLTR is able to capture the signals that lead to accurate feature rankings among training data, potentially correct the noise in the data and use the signals to prioritize features for new patients.

Table 1 also shows that the best performance for the 35 testing patients is better than that for the 163 testing patients (e.g., AH@5 = 1.886 for 35 testing patients vs AH@5=1.429 for 163 testing patients). This indicates that when there are more similar patients for model training, PLTR is able to achieve better performance. However, when there are more testing patients and thus the similarities between training and testing patients are smaller, PLTR achieves more significant improvement compared to the baseline methods (e.g., 1.429/0.356 = 4.01 for 163 testing patients vs 1.886/0.971=1.94 for 35 patients). This indicates that when patient similarities are smaller, PLTR is able to achieve much better improvement over the baseline methods.

Feature Prioritization on Population Level We also investigate which features are frequently prioritized for all the testing patients. We sort all the top-5 ranked features from all the testing patients, weighted by their aggregated ranking positions among the patients, so that features that are frequently ranked high among many patients will be sorted on top. Table 2 lists the top 10 frequently prioritized features by PLTR among the 163 testing patients. Among these 10 features, 8 of them are among the top 10 identified from the ground truth. Sim-

ilarly, for the 35 testing patients, 7 of the top 10 most frequently prioritized features are among the top 10 identified from the ground truth. This indicates the capability of PLTR to find common AD biomarkers on a population level.

Table 2: Top-10 frequent features by PLTR (simU, $\sigma = 1$)

rank	features	p-value	GT
1	ctx-lh-frontal pole	8.67e-20	Y
2	ctx-rh-frontal pole	5.68e-20	Y
3	right-lateral ventricle	4.34e-04	Y
4	ctx-rh-medial orbitofrontal	4.79e-23	Y
5	left-lateral ventricle	1.09e-04	Y
6	ctx-lh-rostral middle frontal	5.12e-21	Y
7	right-choroid plexus	4.41e-05	N
8	ctx-rh-rostral middle frontal	3.68e-20	N
9	ctx-lh-precuneus	3.19e-19	Y
10	non-wm-hypointensities	8.75e-01	Y

The p-value measures whether the feature means are statistically different between controls and patients. Column "GT" indicates if the feature is in ground truth (Y) or not (N). These features are frequently prioritized by PLTR when 163 patients are hold out for testing.

Most of the above top ranked amyloid features are related to AD or its biomarkers. For example, frontal lobe, the region where frontal pole, rostral middle frontal gyrus and medial orbitofrontal cortex are located, shows significantly higher amyloid deposition in AD/MCI patients than in controls [3]. Furthermore, Huang *et al.* [6] report that both frontal lobe and precuneus show significantly higher amyloid deposition in both MCI and AD compared to HC. Additionally, they report the negative correlation between Mini-Mental State Examination (MMSE) score with amyloid deposition in frontal lobe and precuneus, which further validates increased amyloid deposition in these regions of MCI and AD patients.

5.2 Study on Patient-Patient Similarities

Table 3 presents the best performance when a different patient similarity is applied. In this case, the patient similarities are calculated using an RBF kernel ($\sigma = 5$) on the FreeSurfer features of the patients, after the FreeSurfer features are divided by the corresponding feature mean from normal patients. This feature normalization measures how much the FreeSurfer features in patients deviate from those in HCs. This similarity measurement is denoted as simN. 62 patients are hold out for testing, who have at least 10 training patients each with patient similarities higher than 0.65. The feature ranking is done in the same way as in Section 5.1. Table 3 shows that the PLTR substantially outperforms BMTMKL and KRL. Table 3 and Table 1 together demonstrate that regardless of similar functions used to measure patient similarities in FreeSurfer features, PLTR is

Table 3: Overall Performance of PLTR (simN, $\sigma = 5$)

n	method	Parameters					AH@5	WAH@5	AH@10	WAH@10
		α	β	γ	d	λ				
62		0.5	1.0	1.0	10	-	1.371	1.161	**3.129**	**2.295**
	PLTR	0.5	1.0	1.0	20	-	1.387	**1.186**	3.081	2.162
		0.5	1.0	1.0	50	-	**1.403**	1.165	3.113	2.117
	BMTMKL	-	-	-	-	-	0.790	0.670	1.871	1.982
	KRL	-	-	-	-	0.5	0.306	0.299	0.968	1.046

The column "n" corresponds to the number of hold-out testing patients. Best performance under each evaluation metric is in **bold**.

robust in outperforming baseline given that the testing patients have sufficient similar training patients.

6 Conclusions and Discussions

We have proposed an innovative machine learning paradigm enabling precision medicine for AD imaging biomarker prioritization. The paradigm tailors the imaging biomarker discovery process to individual characteristics of a given patient, and has been implemented based on a newly developed learning-to-rank method PLTR. To the best of our knowledge, this learning-to-rank method has never been applied to the AD imaging biomarker studies. It is a paradigm shifting strategy to facilitate precision medicine research in brain imaging study of AD. The PLTR model seamlessly integrates two objectives for joint optimization: pushing up relevant biomarkers and ranking among relevant biomarkers. The empirical study of PLTR has been performed on the ADNI data and yielded promising results to identify and prioritize individual-specific amyloid imaging biomarkers based on the individual's structural MRI data.

Acknowledgements

This work was supported in part by NIH R01 EB022574, R01 LM011360, U19 AG024904, R01 AG019771, and P30 AG010133; NSF IIS 1837964 and 1855501. The complete ADNI Acknowledgement is available at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

1. Alzheimer's Association: 2017 Alzheimer's disease facts and figures (2017)
2. Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Hintsanen, P., Khan, S.A., Mpindi, J.P., et al.: A community effort to assess and improve drug sensitivity prediction algorithms. Nature biotechnology **32**(12), 1202 (2014)

3. Forsberg, A., Engler, H., Almkvist, O., Blomquist, G., Hagman, G., Wall, A., Ringheim, A., Långström, B., Nordberg, A.: Pet imaging of amyloid deposition in patients with mild cognitive impairment. Neurobiology of Aging **29**(10), 1456 – 1465 (2008)
4. He, X., Folkman, L., Borgwardt, K.: Kernelized rank learning for personalized drug recommendation. Bioinformatics **34**(16), 2808–2816 (2018)
5. He, Y., Liu, J., Ning, X.: Drug selection via joint push and learning to rank. IEEE/ACM Transactions on Computational Biology and Bioinformatics pp. 1–1 (2018)
6. Huang, K.L., Lin, K.J., Hsiao, I.T., Kuo, H.C., Hsu, W.C., Chuang, W.L., Kung, M.P., Wey, S.P., Hsieh, C.J., Wai, Y.Y., Yen, T.C., Huang, C.C.: Regional amyloid deposition in amnestic mild cognitive impairment and alzheimer's disease evaluated by [18f]av-45 positron emission tomography in chinese population. PLOS ONE **8**(3), 1–8 (03 2013)
7. Marinescu, R.V., Oxtoby, N.P., et al.: TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. arXiv e-prints arXiv:1805.03909 (May 2018)
8. Ten Kate, M., Barkhof, F., Visser, P.J., Teunissen, C.E., Scheltens, P., van der Flier, W.M., Tijms, B.M.: Amyloid-independent atrophy patterns predict time to progression to dementia in mild cognitive impairment. Alzheimers Res Ther **9**(1), 73 (2017)
9. Weiner, M.W., Veitch, D.P., et al.: The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimers Dement **13**(5), 561–571 (2017)